Comprehensive Data Quality Framework using Great Expectations

# Expanded Implementation Plan & Jira Tasks

# Expanded Implementation Plan

## Phase 1: Assessment & Tooling Setup

* Conduct requirements gathering across stakeholders.
* Install and configure Great Expectations (GX).
* Initialize GX project and setup version control (Git).
* Connect initial data sources for validation.

Deliverables: Requirements Document, GX Setup, Git Repo, Data Source Validation

## Phase 2: Data Profiling

* Profile critical datasets to generate initial expectation suites.
* Review and refine profiling-generated expectations.
* Version control profiling results.

Deliverables: Profiling Reports, Baseline Expectation Suites, Schema Risk Document

## Phase 3: Rule Engine Setup

* Create repository of custom business validation rules.
* Implement dynamic expectation suite generation.
* Group suites by domain for modularity.

Deliverables: Rule Templates, Dynamic Suite Scripts, Domain Organized Suites

## Phase 4: Integration & Automation

* Integrate GX validations into ETL pipelines (Airflow, DBT, etc.).
* Create scheduled validation jobs.
* Setup CI/CD pipelines for data quality runs.

Deliverables: Automated Pipelines, Dynamic Validation Jobs, CI/CD Integration

## Phase 5: Reporting & Alerts

* Build and publish GX Data Docs.
* Configure alerting for validation failures (Slack/Email).
* Build Data Quality Dashboards and auto-ticketing integrations.

Deliverables: Data Docs, Alerts Setup, Dashboards, Ticketing Integration

## Phase 6: Testing, Optimization & Handoff

* Write unit and integration tests for validations.
* Optimize validation performance (batching, Spark).
* Prepare handoff documentation and KT sessions.

Deliverables: Test Reports, Optimized Configs, Documentation, KT Sessions

# Jira Stories

## Assessment & Tooling Setup

* DQ-1: Gather data quality requirements
* DQ-2: Install and configure Great Expectations
* DQ-3: Setup Git for version control
* DQ-4: Connect to initial data sources

## Data Profiling

* DQ-5: Profile critical datasets to generate initial expectation suites
* DQ-6: Review and refine profiling results
* DQ-7: Version control profiling suites

## Rule Engine Setup

* DQ-8: Create repository of custom business rules
* DQ-9: Implement dynamic expectation generation
* DQ-10: Group suites by domain

## Integration & Automation

* DQ-11: Integrate GX validations into ETL pipelines
* DQ-12: Schedule daily validation jobs
* DQ-13: Integrate GX validations into CI/CD pipelines

## Reporting & Alerts

* DQ-14: Generate Data Docs after validation
* DQ-15: Send alerts on validation failures
* DQ-16: Build Data Quality Dashboards
* DQ-17: Setup auto-ticketing for critical failures

## Testing, Optimization & Handoff

* DQ-18: Write unit and integration tests for validation
* DQ-19: Optimize validations for large datasets
* DQ-20: Prepare handoff documentation and KT materials
* DQ-21: Establish continuous improvement loop